CA04/1462

PA 1234375

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

October 08, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: *60/494,781*
FILING DATE: *August 14, 2003*

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

H. L. JACKSON
Certifying Officer

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2)

Docket Number	11515-1	Type a Plus Sign (+) inside this Box -->	+

INVENTOR(S)/APPLICANT(S)

LAST NAME	FIRST NAME	Middle Initial	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Somorjai	Ray	L.	146 Grange St., Headingley, Manitoba R4H 1A8, Canada
Smith	Ian	C.P.	63 Shier Drive, Winnipeg, Manitoba R3R 2H2, Canada

TITLE OF THE INVENTION (280 CHARACTERS MAX.)

Method of Diagnosing Colorectal Adenomas and Cancer Using Infrared Spectroscopy

CORRESPONDENCE ADDRESS

Margaret McKay
National Research Council of Canada
Intellectual Property Services Office, EG-06B, Bldg. M-58
Montreal Road, Ottawa, Ontario, Canada K1A 0R6

STATE	Ontario	ZIP CODE	K1A 0R6	COUNTRY	Canada

ENCLOSED APPLICATION PARTS (*Check all that Apply*)

X	*Specification*	*Number of pages*	*11*			*Small Entity Statement*	
X	*Drawing(s)*	*Number of Sheets*	0			*Other (specify)*	

METHOD OF PAYMENT (*Check One*)

	A check or money order is enclosed to cover the Provisional filing fees	*Provisional Filing Fee Amount ($)*	$ 160.00
X	*The Commissioner is hereby authorized to charge filing fees and any deficiency in the filing fees and credit to our* *Mastercard Account* *Form PTO-2038 is attached*		
X	*We hereby authorize you to deduct any deficiency of the fee stated above or credit to this amount to our Deposit Account 14-0429.*		

The invention was made by an agency of the United States Government or under contract with an agency of the United States Government. X No

Date:

14Aug03

Enclosures
:yw

Respectfully submitted,

Margaret McKay
Patent Agent for Applicant
Regn No. 52,519

X Additional inventors are being named a separately numbered sheets attached hereto.

PROVISIONAL APPLICATION FILING ONLY

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(b)(2)

Docket Number	11515-1	Type a Plus Sign (+) inside this Box -->	+

INVENTOR(S)/APPLICANT(S)			
LAST NAME	FIRST NAME	Middle Initial	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Meltzer	John	C.	463 Oxford St. Winnipeg, Manitoba R3M 3J1, Canada
Nikouline,	Alexandre		35 Pine Bluff Rd., Winnipeg, Manitoba R2J 2N6, Canada
Dolenko,	Brion		490 Kingsford Avenue, Winnipeg, Manitoba R2G 0K2, Canada

METHOD OF DIAGNOSING COLORECTAL ADENOMAS AND CANCER USING INFRARED SPECTROSCOPY

This invention relates to a method of detecting colorectal adenomas and cancer, and in particular to a method of detecting such adenomas and cancer using near infrared spectroscopy.

Colorectal cancer is one of the most common cancers in the U.S.A. and 105,000 people are expected to develop this disease in 2003; it is also projected that 57,000 will die of this in the U.S.A in 2003. The lifetime risk that an individual in North America will develop colorectal cancer is believed to be about 5 - 6 %. Symptoms associated with colorectal cancer, including blood in the stool, anemia, abdominal pain and alteration of bowel habits often become apparent only when the disease has advanced significantly. It is well known that prognosis for a patient depends largely on the stage of the disease at the time of diagnosis. In fact, whereas the five-year survival for a patient whose colorectal cancer is detected at an early stage is 92%, survival decreases to about 60% in patients with regional spread, and to about 6% in those with distant metastases. Accordingly, it is important to detect the precursor adenomas and cancer as early as possible to increase the chances of successful therapeutic intervention.

A screening technique preferably provides high sensitivity and specificity, low cost, safety and simplicity. Currently, digital rectal examination (DRE), fecal occult blood test (FOBT), barium enema and direct colon visualization (sigmoidoscopy and colonoscopy) screening techniques are employed.

DRE involves examining the rectum using a finger. This method detects cancers that can be palpated and are within reach of the finger. A negative DRE provides little reassurance that a patient is free of cancer, because fewer than 10% of colorectal cancers can be palpated by the examining finger.

FOBT detects hidden blood in the stool by chemical means. Although the least expensive and the simplest, the FOBT method has low sensitivity, moderate specificity and is usually not good for early detection. According to available data, a major drawback of this technique is that more than half of the cancers discovered by this method followed by x-ray or endoscopy are usually beyond the limit of early staging. A false positive rate of 10-12% is expected when the patients tested are on an unrestricted diet. Estimates of the positive predictive value range from 2.2 to 50%. The guaiac tests have a very low sensitivity, generally around 50%. The use of FOBT is based on the assumption that colorectal cancers are associated with bleeding. However, it appears that some colorectal cancers bleed intermittently and others not at all.

A barium enema involves an x-ray of the bowel using a contrast agent. The enema can be a single or double contrast. The main radiologic signs of malignancy include muscosal disruption, abrupt cut-off and shouldering and localized lesions with sharp demarcations from uninvolved areas. The estimated sensitivity of double contrast barium enema for cancer and large polyps is only about 65-75% and even lower for small adenomas. Despite its better diagnostic yield, double contrast barium enema has a false-negative rate of 2-18%. Moreover, the method involves exposure to radiation, the repeated use of which may not be safe. Perforation from barium enema is extremely uncommon, but when it happens it is can be fatal or lead to serious long term problems as a result of barium spillage into the abdominal cavity.

A variety of instruments (collectively called endoscopes) are generally used for examining the bowel. Endoscopes can be rigid or flexible with varying lengths. Flexible sigmoidoscopes are 60 cm long. A colonoscope is a 130 - 160 cm flexible viewing instrument for examining the entire colon. Biopsies are taken from suspicious looking areas while viewing the colon through the endoscope. The flexible sigmoidoscopy examination is limited to the left side of the colon and rectum. At least 1/3 of neoplastic tumors are believed to occur in areas proximal to the splenic flexure that are inaccessible by sigmoidoscopy. Colonoscopy has a high sensitivity, and remains the gold standard for visualization of the colon and the detection of neoplastic abnormalities. However, it is invasive, quite expensive, and exposes the subject to risks of bowel perforation.

There are a number of currently available methods for detecting cancer in its early stages. Biophysical methods such as conventional X-rays, nuclear medicine, rectilinear scanners, ultrasound, CAT and MRI all play an important role in early detection and treatment of cancer. Clinical laboratory testing for tumor markers can also be used as an aid in early cancer detection. Tumor marker tests measure either tumor-associated antigens or other substances present in cancer patients which aid in diagnosis, staging, disease progression, monitoring response to therapy and detection of recurrent disease. Unfortunately, most tumor marker tests do not possess sufficient specificity to be used as screening tools in a cost-effective manner. Even highly specific tests often suffer from poor predictive value, because the prevalence of a particular cancer is relatively low in the general population. The majority of available tumor marker tests are not useful in diagnosing cancer in symptomatic patients because elevated levels of markers are also seen in a variety of benign diseases. The main clinical value of tumor markers is in tumor staging, monitoring therapeutic responses, predicting patient outcomes and detecting recurrence of cancer.

Magnetic resonance spectroscopy (MRS) is a technique that has the potential to detect small and early biochemical changes associated with disease processes, and has been proven to be useful in the study of tissue biopsies from cancer patients. It is particularly useful to detect small, mobile chemical species in a given biological sample that are of diagnostic interest. Obtaining tissue biopsies for such an examination, however, usually involves an invasive procedure. C.L. Lean et al (Magn. Reson Med 20:306-311, 1991; Biochemistry 3:11095-11105, 1992 and Magn Reson Med 30:525-533, 1992) describe the use of magnetic resonance spectroscopy to examine colon cells and tissue specimens. Bezabeh et al (Bezabeh T, I. C. P Smith, B Levin, C. Berstein, and C Johnson. WO 00/71997 A1 Method of Diagnosing Colorectal Adenomas and Cancer Using Proton Magnetic Resonance Spectroscopy; Bezabeh T, R Somorjai, B Levin, I. C. P Smith, C. Berstein, and C Johnson. WO 02/12879 A Method of Diagnosing Colorectal Adenomas and Cancer Using Proton Magnetic Resonance Spectroscopy.) demonstrate a method to diagnose colorectal adenomas and cancer using MRS on stool samples. WO 00/71997 A1 teaches the use of 2 dimensional MRS to delineate specific chemical species such as fucose which may be indicative of colon cancer and polyps. WO 02/12879 teaches the use of 1 dimensional MRS and a classification-based strategy to differentiate between diseased and non-diseased patients based on their stool. Although useful, MRS requires more expensive equipment.

Infrared spectroscopy (IRS) is a lower-cost technique that has the potential to detect small and early biochemical changes associated with disease processes.

The infrared (IR) region of light is in between the visible and microwave portions of the electromagnetic spectrum. The IR spectral region ranges from 780 to 25,000nm (12800cm^{-1} to 400cm^{-1}) and is commonly subdivided into further regions including the near-IR (4000-12800 cm^{-1}) and mid-IR (400-4000^{-1}). IR spectroscopy measures the absorption of infrared radiation by chemical bonds. Therefore IR spectra contain the basic vibrational fingerprints of all molecules examined in a particular sample and this information can provide insight on nature of the chemical bonds, the structure and the microenvironment of the sample being studied.

Fragments of molecules, known as functional groups, tend to absorb IR radiation in the same frequency regardless of the structure of the rest of the molecule containing the functional group. For example absorptions between 1620-1680^{-1} are usually attributed to the amide I vibration of proteins while absorptions at 1080 and 1240 cm^{-1} are attributed to the

PO2- symmetric and asymmetric stretching vibrations of DNA phosphodiester groups .

IR spectroscopy can be used to study substances such as carbohydrates, proteins, lipids and DNA in isolation or as part of complex biological samples. Such biological samples include tissues (for example whole tissues in vivo or ex-vivo, tissue slices, histological sections and cell suspensions) and fluids (for example urine, blood, amniotic fluid), even if the fluids are first dried onto an IR- compatible substrate.

IR spectroscopy can be used in various modalities to study biological samples including transmission, attenuated total reflectance, diffuse reflectance and Raman Spectroscopy. Data processing techniques such as spectral subtraction, spectral derivatives, deconvolution, multivariate analysis (such as linear discriminate analysis and partial least squares regression) and unsupervised methods (such as principal components analysis and various clustering techniques) are then used to analyze the complex IR spectroscopic data.

IRS can be performed with relatively inexpensive equipment. It has been used for clinical chemistry applications with IR-transparent substrates such as Barium Fluoride and with substrates that have limited IR-transparency such as glass demonstrating its utility and its potential as a cost-effective modality for mass-screening.

IRS has been proven to be useful in the study of tissue biopsies from cancer patients including tissue samples from patients with colon cancer. Human colon adenocarcinoma cell lines display infrared spectroscopic features of malignant colon tissues.

These findings have been extended to the in-vivo and ex-vivo analysis of colon polyps by near infrared Raman spectroscopy and multivariate statistical techniques.

IRS analysis has also been used to screen for colon cancer by the fecal occult blood test by optically detecting the presence of blood in smeared stool samples and IRS has been used to assess the location of gastric bleeding based on the spectroscopic analysis of centrifuged stool samples by means of an artificial neural net.

IRS has also been used on stool to assess nutrient uptake by measuring fecal polyethylene glycol, fecal fat levels, etc., all by measuring known chemicals at specific peaks.

There is provided herein a relatively simple, non-invasive method of detecting colorectal adenomas and cancer that meets the above- defined criteria of high sensitivity and specificity, low cost and safety.

Accordingly, the invention provides a method of detecting the presence of colorectal adenomas and colorectal cancer in a patient comprising the steps of subjecting a stool sample from the patient to infrared spectroscopy; and comparing the resulting spectrum with the infrared spectra of stool from non-cancerous subjects, with observed differences in spectra being indicative of cancer or clinically significant adenomas.

In one embodiment of the method, the stool sample is mixed with a buffer to produce a suspension of stool sample, the suspension is centrifuged to yield a supernatant sample, the supernatant sample is subjected to infrared spectroscopy, and the resulting spectrum is compared with infrared spectra of stool from non-cancerous subjects.

The performing of spectral analysis on human stool offers a significant advantage over other methods, because the collection of the specimen is non-invasive and presents no risk to the patient. Moreover, no special processing of the sample is required prior to analysis.

METHOD

Subjects who were scheduled for colonoscopy or surgery were recruited to donate a single sample of stool. Table 1 provides a breakdown of the cases used.

Table 1

(Breakdown of subjects recruited)

Cases	
Colorectal cancer or Adenomatous Polyps	100
Normal	100

The group referred to as "Normal" includes some subjects with colonic conditions/abnormalities that are non-neoplastic. Examples Include diverticulosis, hyperplastic polyps and internal hemorrhoids. Specimens from subjects with inflammatory bowel disease were not included in the analysis.

Stool samples were collected at the University of Texas M.D. Anderson Cancer Center; University of Manitoba, Health Sciences Centre; University of Chicago; and University of Toronto, Mount Sinai Hospital. Subjects were instructed to collect their bowel movements prior to their colonic preparations. The samples were kept frozen in the patients' refrigerators for an average of 24-48 hours prior to their delivery to the hospital in small ice chests (mailers). They were then stored in a -70 degree C freezer until being shipped "blinded", on dry ice, to the National Research Council Institute for Biodiagnostics, Winnipeg, Canada. All samples were shipped in dry ice and kept frozen at –70 degrees C until the time of the experiment. There was no significant difference in the lengths of time for which the samples were kept frozen. All samples were randomly assigned a code number that was not traceable to the original sample.

SAMPLE PREPARATION

For IRS experiments, samples were thawed and a portion of the sample was then taken and suspended in saline. The suspension was then gently vortexed, and replicate dry films were prepared by depositing about 5 μl of the suspension on an infrared-transparent (barium fluoride-BaF_2) window and drying it down quickly under mild vacuum as a thin circular film of 2 – 3 mm diameter. The remaining sample was then centrifuged and replicate films were prepared by drying 15 μl aliquots onto BaF_2 windows. After measurements, the materials in the windows were washed out with 70% alcohol and water and the waste was stored at the biohazard container. During preparation, the operator wore gloves throughout the procedures to avoid any potential contamination.

IRS EXPERIMENTS

For each sample, single beam IR spectra were ratioed against the spectrum of a blank barium fluoride window and converted to absorbance units. All spectra were acquired using a Bio-Rad FTS-60 IR spectrometer equipped with a nitrogen cooled mercury cadmium telluride detector, set at a nominal resolution of 2 cm^{-1} and an encoding interval of one wavenumber. For each spectrum, 256 interferograms were co-added and apodized with a triangular smoothing function before Fourier transformation. Each sample was run twice, resulting in two replicate spectra. This made it possible to check for inconsistencies in the IR processing.

DATA PROCESSING (I)-

A region consisting of 1,608 data points from each spectrum was used for the analysis. This covered most of the mid-IR range, from 900cm^{-1} to 4000cm^{-1}. Each spectrum was then normalized by dividing every data point by the total spectral area. Depending on the data set, it may be advantageous to perform further processing according to methods known to those skilled in the art in light of the disclosure herein. By taking first derivatives, offsets between the spectra were eliminated. The first derivative used simply replaced each data point by the difference between it and the adjacent data point. Performing this operation a second time yielded a second derivative, which eliminated any differences in baseline slopes between spectra. After either derivative is taken, or even if no derivative is used, one may rank order the spectral intensities, replacing the smallest intensity by 1, second smallest by 2, and so on up to the largest intensity, replaced by N, where N is the number of intensity values. This can help in making robust any methods to discriminate between the classes of data, by keeping all the data within the same bounds. A spectrum that originally contained a very large peak did not appear as great an outlier to a classifier after rank ordering.

The statistical classification strategy used has been developed specifically to deal with the discrimination of spectra of biomedical origin. The strategy comprises three stages. The first stage is a preprocessing step, found to be preferred for reliable classification. It consists of selecting from the spectra a few maximally discriminatory subregions, using an optimal region selection (ORS) algorithm, based on a genetic algorithm (GA)-driven optimization method (Nikulin A E, Dolenko B, Bezabeh T, Somorjai R L, NMR in Biomedicine 11, 209-217 (1998), Near-optimal region selection for feature space reduction: Novel preprocessing methods for classifying MR spectra; Bezabeh, T. et al, The use of ^{1}H Magnetic Resonance Spectroscopy in Inflammatory Bowel Disease: Distinguishing Ulcerative Colotis From Crohn's Disease, Am. J. Gastroenterol 2001, 96: 442-448; Somorjai, R.L. et al, Distinguishing Normal from Rejecting Renal Allographs: Application of a Three-Stage Classification Strategy to MR and IR Spectra of Urine, Vibrational Spectroscopy 28 (1) 97-102 (2002) and Lean, C.L. et al, Accurate Diagnosis and Prognosis of Human Cancers by Proton MRS and a Three-stage Classification Strategy, Annual Reports on NMR Spectroscopy 2002, 48: 71-111). For reliability of classification, the number of these subregions are preferably an order of magnitude smaller than the number of samples to be classified.

The ORS algorithm was run several times, each using different starting points on each of several different random splits of the data. For each split, roughly 2/3rds of the samples (two replicate spectra each) were selected for the training set (used to construct the classifier), and the remainder were used as a test set (to estimate the classifier's prediction accuracy on new samples). This method of several random splits is preferable to using just one training and test set, as inevitably some training sets will be more representative than others of the entire possible data space. Classifiers trained using these data sets will show higher accuracies on the test samples, and will also have better accuracies on new samples.

Generally 2/3rds of the samples in the smallest class are selected for training, and then an equal number of samples (may be a smaller percentage) in the larger class are selected. This eliminates any significant imbalances in the number of samples for each class; a large class cannot overwhelm a smaller one (and make it more difficult to classify). Nevertheless, if one class still proves much more difficult to classify than the other, that class can be given more weight, making it more important in scoring the subregions.

Due to the non-exhaustive nature of the ORS algorithm, it is entirely possible that certain subregions from one data split, when combined with subregions from another data split, will yield higher classification accuracies than when used alone. Investigators may collect a large number of promising subregions, and then exhaustively search through all possible subsets for a small number of subregions that still yields good classification accuracy. As already stated, the number of subregions should be kept small for reliability of classification.

Once a set of optimal subregions has been found, the second stage involves computing an ultimate classifier based on those regions. To avoid the overly optimistic classification results that a straight resubstitution approach would give, the inventors have developed a cross-validation method, using a bootstrap methodology. The bootstrap method repeatedly partitions (with replacement) the data into many approximately equal sized random training and test subsets. For each of the random training subsets an optimal classifier is found, and its accuracy is validated on the random test subset. The process is repeated a number of times, usually 10000. The ultimate classifier is a weighted average of the classifier coefficients of the 10,000 individual component classifiers. This approach effectively uses all n samples.

A standard multivariate statistical method, Linear Discriminant Analysis (LDA) is preferred choice for all classifiers, at all stages, because of its speed and robustness. The concept of crispness of a classifier is also used because the inventors' classifiers produce class probabilities. As used herein a 2-class classification of a sample is considered crisp if the class assignment probability for that sample is >75%. This crispness is used in the weighting of the classifier coefficients at the bootstrap stage – the weight includes the percentage of samples crisply classified, and Cohen's Kappa ($\kappa(0.5,0)$), the latter being a measure that indicates the goodness of classification above chance. Similar measures are also used when scoring classifiers at the ORS stage. Generally subregions producing classifiers with high crispness and Cohen's Kappa values on the test sets are chosen as the optimal ones. Optionally a penalty function can be used to help minimize the difference in accuracies between the normal and cancer classes.

For difficult classification problems, a third stage consists of combining the outcomes of several classifiers via aggregation methods (computerized consensus diagnosis, CCD) into an overall classifier that is more reliable and accurate than the individual classifiers.

The particular classifier aggregation used by the inventors is one of the variants of Wolpert's Stacked Generalizer (WSG) (Wolpert D H, Stacked Generalization. Neural Networks 5, 241-259 (1992)). The version of WSG used takes the output class probabilities obtained by the individual classifiers as input features to the ultimate classifier. For 2-class problems the number of features is 1 per classifier (with K independent classifiers this gives K probabilities as input features). The overall classification quality is generally higher. The crispness of the classifier is greater. This is important in a clinical environment because fewer patients will have to be re-examined.

10 regions obtained from an earlier classifier development were used to produce the results reported by using 1st derivatives, 1st derivatives rank ordered, 2nd derivatives, 2nd derivatives rank ordered (4 different classifiers). The probabilities produced by these 4 classifiers were then combined by stacked generalization. 10 random splits of the data were made.

Results are (based on 67 normal samples+ 67 cancer samples in the training sets (x2 replicate spectra for each sample) and 33 normal + 33 cancer samples in the validation sets (x2 spectra for each)) 81.2+/- 1.4 % for the average of the 10 training sets (TSs) and 80.7 +/- 2.2% for the average of the 10 validation sets (VSs). The crisp results are 86.0 +/- 0.8% (78.3 +/- 2.8% crisp) (TSs), 84.6 +/- 3.6% (90.1 +/- 5.6% crisp) (VSs).

In an embodiment of the invention there is provided a method of diagnosing cancer of the colon or rectum, said method comprising:

(a) obtaining a sample of excrement which has passed through the colon or rectum;

(b) subjecting the sample, or a portion thereof, to infrared spectroscopy; and

(c) comparing the resulting spectrum with the infrared spectra of similar samples from (i) subjects suffering from colon or rectal cancer and (ii) subjects not suffering from such cancers.

In an embodiment of the invention there is provided a method of analyzing spectral data, said method comprising:

a) normalizing the data points;
b) computing derivatives of the spectra;
c) rank ordering the spectral intensities or their derivatives; and,
d) obtaining data points from the spectrum covering a spectral region of interest.

In some instances the data points are normalized by dividing each data point by the total area of the spectrum.

In an embodiment of the invention there is provided a statistical classification method suitable for use in the discrimination of spectra of biomedical origin. The method comprises:

(a) selecting from subject spectra a number of sub-regions which tend to differ strongly between control spectra obtained from subjects having different medical conditions of interest;

(b) producing an ultimate classifier based on the regions selected in step (a); and,

(c) combining the outcomes of at least some of the classifiers created in step (b) using an aggregation method to obtain an overall classifier.

In step (a) the selection is preferably made using an optimal region selection (ORS) algorithm. The ORC algorithm is preferably based on a genetic algorithm (GA)-driven optimization method. The ORS algorithm is preferably run several times, each using different starting points on each of several different random splits of the data.

In some instances it is preferable to select from the spectra (in step (a)) a number of sub-regions, which number is about an order of magnitude smaller than the number of samples to be classified.

In step (b) a cross-validation method is preferably employed in computing the ultimate classifier. The cross-validation method preferably partitions the data into plurality of substantially equal sized random training and test subsets. This process is repeated at least 1000, and preferably about 10,000 times, each repetition generating an individual component classifier. The ultimate classifier is then calculated as a weighted average of the classifier coefficients of the 10,000 individual component classifiers.

In some instances Linear Discriminant Analysis ("LDA") is employed in developing the various classifiers. However, LDA may not be the optimal classifier in all circumstances. The method therefore contemplates the use of other classification methods which may be better suited for use with a particular dataset. In step (c), any suitable aggregation method (or combination of methods) may be employed, including without limitation computerized consensus diagnosis.

ABSTRACT OF THE DISCLOSURE

Infrared spectroscopy of human stool can be used as a non-invasive method of detecting the presence of colorectal cancer and/or clinically significant adenomas. The spectrum of a patient's stool is compared with that of stool from non-cancerous subjects, observed differences in spectra being indicative of cancer and/or clinically significant adenomas. In a preferred method, the stool sample is mixed with a buffer, the resulting suspension is centrifuged and the supernatant is subjected to infrared spectroscopy. The spectra are then classified using a three-stage classification strategy.